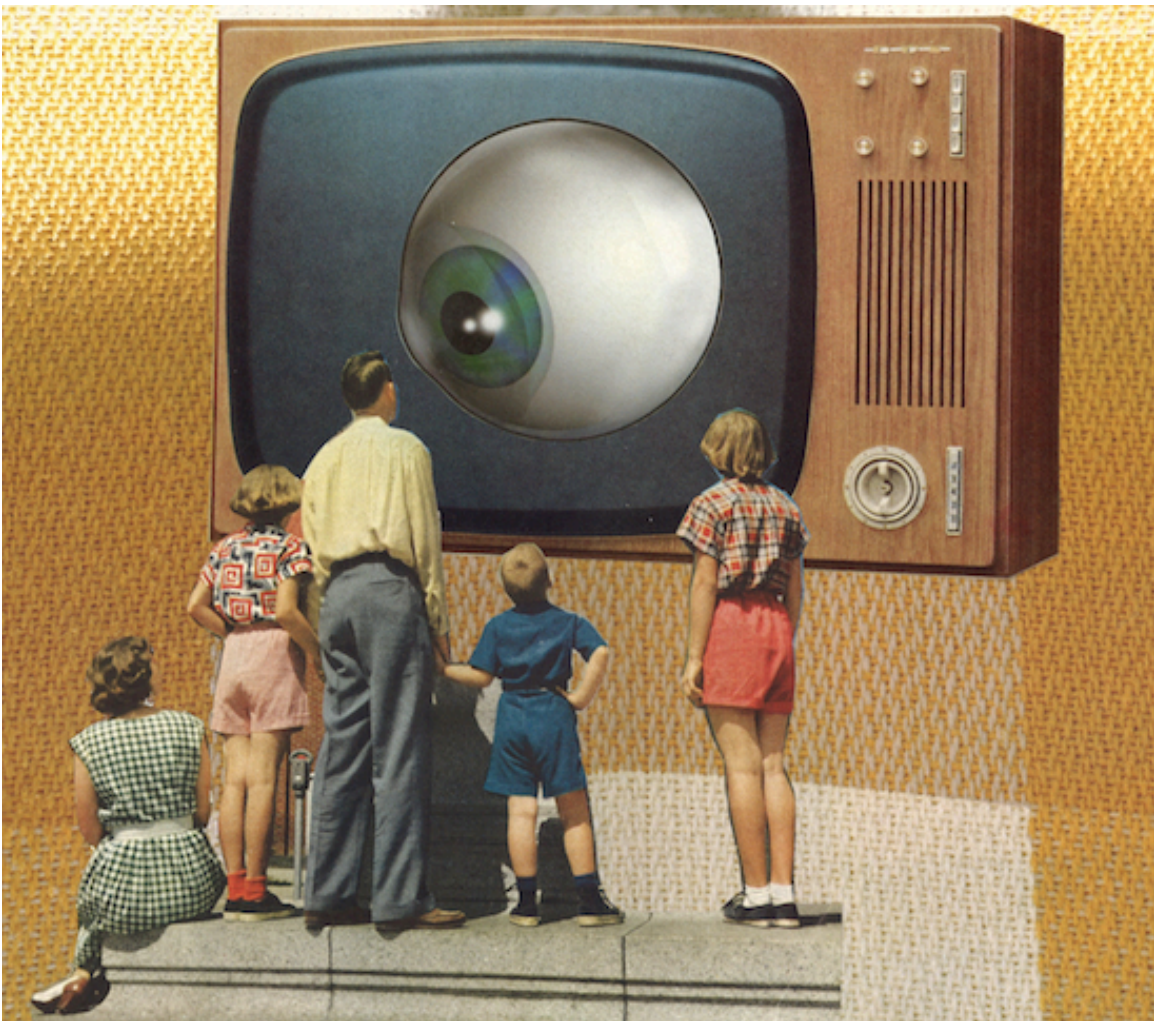


AI = Apocalypse Imminent?

or

How Many Ways Could We Die By AI?



Phoenix X Eeyore

AI = Apocalypse Imminent?

Imagine

I previously thought we had more time. At least a few years. But now it feels like I erred.

Imagine that you faced immediate and total annihilation of you and everyone you loved. But imagine that you could go back in time a few months. What would you do?¹

Would you choose to simply enjoy those few months, to indulge in each breath, to cherish each vital relationship in your life, to help someone left outside in the cold? Or would you focus all of your energy on preventing the impending annihilation if you believed you could still do something about it? What if you could do both at the same time? If human societies act as a single body, we could do both. The two options may, in fact, go hand in hand.

Massive Change

I believe we stand on the brink of such total annihilation in various ways... and that a solution of radical, mass, uncompromising, and nonviolent action by a relatively small group of people could address all of those ways (whether averting nuclear war, ecological catastrophe, or tech-spurred apocalypse). I spend this essay addressing one of them: AI.

Perhaps, in part, most of us have not yet felt compelled to take immediate and radical change in our relationships to life and each other because we have little idea about what exists in the world today—much less the tech-tsunami of chaos on the horizon. Also, most of the world with wealth and power live in complete detachment and alienation from the rest of the world who already live in an apocalypse of toxic environments, insecurity, and deprivation from vital basics such as decent water, housing, and food.

So, to begin with, we can list what tech exists in the world today:

- We live in a world in which minuscule robots can swim in human blood veins to perform tasks such as delivering medicine and/or toxins.²
- We live in a world in which thousands of humans have already voluntarily implanted computer chips under their skin³ (and also a world wherein many countries such as the UK, France, Sweden, Israel, Japan, Australia, and Spain have

¹ This does not mean that *right now* we necessarily only have a few months left. But we will someday.

² Leah Crane, “Tiny robots swim the front crawl through your veins.” *New Scientist* 24 July 2017. <https://www.newscientist.com/article/2141595-tiny-robots-swim-the-front-crawl-through-your-veins/>

³ See Dan Lohrmann, “Chip Implants: Opportunities, Concerns and What Could Be Next.” *GovTech*, 16 January 2022. <https://www.govtech.com/blogs/lohrmann-on-cybersecurity/chip-implants-opportunities-concerns-and-what-could-be-next> ; Katherine Latham, “The microchip implants that let you pay with your hand.” *BBC* 11 April 2022. <https://www.bbc.com/news/business-61008730>

mandated implanting trackable chips in animals such as horses, dogs, and cats⁴). Already in 2017, *USA Today* wrote: “You will get chipped—eventually.”⁵

- We live in a world in which brain implants can “spy” on neural activity.⁶
- We live in a world in which governments turn animals into cyborg spies.⁷
- We live in a world in which scientists create and control synthetically produced biological organisms called xenobots.⁸
- We live in a world in which, already in 2014, 3D printers manufactured guns and could self-replicate.⁹ An off-grid 3D maker shop in the desert, powered by the sun produced glass from sand.¹⁰
- We live in a world in which the second richest person on Earth and one of the most cited living philosophers both believe that we live in a computer simulation.¹¹
- We live in a world in which some researchers regard an AI program as “sentient” going so far as to hire a lawyer to defend its rights¹² and a country (Saudi Arabia) granted a robot (Sophia) citizenship.¹³
- We live in a world in which some places (such as Singapore) used “robot dogs” to enforce social distancing during the pandemic.¹⁴ Some companies specialize in constructing robots for military use.¹⁵

⁴ Ross Kelly, “Compulsory cat microchipping gains international momentum,” *Veterinary Information Network*, 2019. <https://news.vin.com/default.aspx?pid=210&id=9359088&f5=1>

⁵ Jefferson Graham, “You will get chipped—eventually.” *USA Today* 2017. <https://eu.usatoday.com/story/tech/2017/08/09/you-get-chipped-eventually/547336001/>

⁶ Elizabeth Gibney, “Injectable brain implant spies on individual neurons,” *Nature* 8 June 2015. <https://www.nature.com/articles/522137a>

⁷ Sarah Laskow, “The government is engineering cyborg spy animals (really),” *Grist* 15 March 2013.

⁸ Joshua Sokol, “Meet the Xenobots, Virtual Creatures Brought to Life,” *New York Times*. 3 April 2020. <https://www.nytimes.com/2020/04/03/science/xenobots-robots-frogs-xenopus.html>

“A xenobot lives for only about a week, feeding on the small platelets of yolk that fill each of its cells and would normally fuel embryonic development. Because its building blocks are living cells, the entity can heal from injury, even after being torn almost in half. But what it does during its short life is decreed not by the ineffable frogginess etched into its DNA—which has not been genetically modified—but by its physical shape. [...] Xenobots with a fork- or snowplow-like appendage in the front can sweep up loose particles (in a petri dish) overnight, depositing them in a pile. Some use legs, of a sort, to shuffle around on the floor of the dish. Others swim [...] All of which makes xenobots amazing and maybe slightly unsettling — golems dreamed in silicon and then written into flesh” (Sokol 2020).

⁹ Neil Mohr, “The 3D printers that print themselves: how RepRap will change the world.” *TechRadar* 1 July 2014. <https://www.techradar.com/news/world-of-tech/future-tech/the-3d-printers-that-print-themselves-how-reprap-will-change-the-world-1255490>

¹⁰ Lori Zimmer, “Solar-Powered 3D Printer Turns Desert Sand Into Glass Bowls.” *Inhabitat*, 19 March 2014. <https://inhabitat.com/the-solar-powered-sinter-3d-printer-turns-desert-sand-into-glass/>

¹¹ Alex Hern, “Elon Musk: ‘Chances are we’re all living in a simulation’,” *The Guardian* 2 June 2016.

¹² Nitasha Tiku, “The Google engineer who thinks the company’s AI has come to life,” *Washington Post* 11 June 2022. <http://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>

¹³ Dom Galeon, “Saudi Arabia Made a Robot a Citizen. Now, She’s Calling For Women’s Rights,” *Futurism* 15 December 2017. <https://futurism.com/saudi-arabia-made-robot-citizen-calling-womens-rights>

¹⁴ BBC, “Coronavirus: Robot dog enforces social distancing in Singapore park,” *BBC* 11 May 2020. <https://www.bbc.com/news/av/technology-52619568>

¹⁵ Luke Dormehl, “Meet Ghost Robotics, the Boston Dynamics of combat bots,” *DigitalTrends* 7 December 2020. <https://www.digitaltrends.com/cool-tech/ghost-robotics-military-bots/>

- We live in a world in which semi-autonomous AI drones made in Turkey, Iran, Russia, and the United States engage in combat as of this writing in Ukraine.¹⁶ At least two countries previously used AI to remotely kill people in other countries.¹⁷
- We live in a world in which industries have wiped out natural habitats and filled environments with pesticides and pollutants such that most insect populations have died out in less than 30 years leaving us all on the brink of insect apocalypse.¹⁸ Meanwhile, corporations develop robotic “bees” to replace the real ones.¹⁹
- We live in a world in which computers have long since acquired the capacity to beat any living human in chess but can now win an art contest with digital painting created in seconds.²⁰ AI programs write essays, screenplays, and news articles indistinguishable from humans.²¹ DeepFake AI programs “revive” the dead from a single photograph and have scammed thousands of people with “stolen” voices.²²
- Finally, even AI researchers see a significant risk that AI will lead to “human extinction or similarly permanent and severe disempowerment of the human species”.²³ In 2015, a group of AI researchers estimated a 50% probability that AI superintelligence will arrive by 2050.²⁴ Development has progressed rapidly since

¹⁶ Morgan Meaker, “Ukraine’s War Brings Autonomous Weapons to the Front Lines,” *Wired* 24 February 2023. <https://www.wired.com/story/ukraine-war-autonomous-weapons-frontlines/> ; Stuart Russell, “AI weapons: Russia’s war in Ukraine shows why the world must enact a ban,” *Nature* 21 February 2023. <https://www.nature.com/articles/d41586-023-00511-5>

¹⁷ Ronen Bergman and Farnaz Fassihi, “The Scientist and the AI-Assisted, Remote-Control Killing Machine,” *New York Times* 18 September 2021. <https://www.nytimes.com/2021/09/18/world/middleeast/iran-nuclear-fakhri-zadeh-assassination-israel.html> ; Robert F. Trager, “Killer Robots Are Here—and We Need to Regulate Them,” *Foreign Policy* 11 May 2022. <https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/>

¹⁸ Caspar A. Hallmann et al., “More than 75 percent decline over 27 years in total flying insect biomass in protected areas,” *PLoS One* 12, no. 10, 2017: 1-21. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0185809>

¹⁹ Daniel Boffey, “Robotic bees could pollinate plants in case of insect apocalypse,” *The Guardian* 9 October 2018. <https://www.theguardian.com/environment/2018/oct/09/robotic-bees-could-pollinate-plants-in-case-of-insect-apocalypse>

²⁰ In 2022, Jason Allen used Midjourney to win an art contest merely by writing instructions for the program to paint from. See Kevin Roose, “An AI-Generated Picture Won an Art Prize,” *New York Times*. 2 September 2022. <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>

²¹ Kevin Roose, “We Need to Talk About How Good AI Is Getting,” *New York Times*, 24 April 2022. <https://www.nytimes.com/2022/08/24/technology/ai-technology-progress.html>

²² Karen Hao and Douglas Heaven, “The year deepfakes went mainstream,” *MIT Technology Review* 24 December 2020; Ashley Belanger, “Thousands scammed by AI voices mimicking loved ones in emergencies,” *Ars Technica* 6 March 2023. <https://arstechnica.com/tech-policy/2023/03/rising-scams-use-ai-to-mimic-voices-of-loved-ones-in-financial-distress/> ; Animated Friedrich Nietzsche [AI DeepFake] https://www.youtube.com/watch?v=TEUvN_gsflY ; Animated Anne Frank [AI DeepFake] https://www.youtube.com/watch?v=_ZatpONJMhA ; Animated Frederick Douglass [AI DeepFake] <https://www.youtube.com/watch?v=2FNKE1EsT40>

²³ AI Impacts, “2022 Expert Survey on Progress in AI,” 2022. <https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/> ; Ezra Klein, “This Changes Everything,” *New York Times* 12 March 2023. <https://www.nytimes.com/2023/03/12/opinion/chatbots-artificial-intelligence-future-weirdness.html>

²⁴ Raffi Khatchadourian, “The Doomsday Invention,” *New Yorker* 23 November 2015. <https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom>

then. What happens when AI programs can produce, code, and develop new improved AI programs better and faster than humans?

Much—or all—of this may sound like science fiction. It may sound surreal or unreal but we live in this world right now. We may see it but do we have the brakes to slow down?

The current rate of technological development—especially AI which, in itself, functions to a large degree like a technological accelerant—seems to rocket us past any capacity to keep up—much less stop it. Even without AGI or “superintelligence,” innovations such as those outlined above, show the potential for autonomous AI bots that could replicate, fix, and fabricate their own parts independent of human intervention. Without immediate, radical, massive change in our lives, the massive changes in material technology can drown us. Yet, as already noted, most people don’t even know what world we live in—much less what lies on the immediate horizon.

Do you have a right to know in which reality you live?

Do you have a right to unmodified existence? Does anyone have a right to live without a chip implant? Does anyone have a right to know that no one will force their children or grandchildren to get a chip implant?

Does anyone own their genetic material or can a corporation own them (and manipulate them)? Can a corporation change one’s DNA? Can the government force people to change any of their DNA (or animals’ DNA) deemed “unsuitable” for society?

Does a child have a right to grow up outside of social media?

Does anyone have the right to not have their image or voice captured on video or recording or, if not, have the right to not have those images or audio disseminated on the Internet or, if not, have the right to not have their voice or image duplicated by AI?

Do we have a right to live in a world without AI? After all, we would not need AI to distribute resources equitably. Should we not sooner need rights to basic needs for all people rather than the forced exposure to an AI-influenced world without our consent?

Who determines “rights” and the norms that inform them? What role do corporations, ad culture, and market pressures play in creating—and enforcing— laws and rights?

These questions may seem overwhelming because we don’t even have them on the table for discussion and, with rapid tech developments coupled with a sea of intoxicants (including cell phones) and economic hardship, most people do not have the time, peace of mind, or local community bonds to organize against it. Far more pressing issues seem all too evident: putting food on the table. Keeping one’s job or getting one. Not getting evicted. Keeping one’s children safe. Not going batshit crazy from stress.

AI, in this context, seems like a solution. People can turn to AI for therapy. To help them communicate with their children, partner, or boss. To help work go quicker or more effectively. People can use AI for endless distraction from the turmoil and pressure of daily life. And, all the while, AI systems assimilate user information and massive Internet or cell phone data to improve its ability to provide more targeted ads and increase its own proficiency at imitating humans until its capacity exceeds that of humans in all key industries.

This essay presents some background on both skepticism toward the possibility of AGI and pro-AI alarmism. It responds with an AI-critical list of actual and likely risks.

AI and its (dis)Contents

So, facing possible futures of incalculable benefits and risks, the experts are surely doing everything possible to ensure the best outcome, right? Wrong. If a superior alien civilisation sent us a message saying, “We’ll arrive in a few decades,” would we just reply, “OK, call us when you get here—we’ll leave the lights on?” Probably not—but this is more or less what is happening with AI.

- Stephen Hawking

Have you ever seen a movie where the machines start thinking for themselves that ends well? Every time I went out to Silicon Valley during the campaign, I came home more alarmed about this. My staff lived in fear that I’d start talking about ‘the rise of the robots’ in some Iowa town hall. Maybe I should have.

- Hillary Clinton

AI, as a term and a field of study, stretches back to the work of three mathematicians and an IBM programmer in 1955 who wrote “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”:

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.²⁵

Thus, these four dudes, John McCarthy, Marvin Minsky, Claude Shannon (the mathematicians), and Nathaniel Rochester (the IBM guy) set the stage for decades of AI research to come. Some of their points and concern remain relevant today:

[T]he major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

How can a set of (hypothetical) neurons be arranged so as to form concepts.

Probably a truly intelligent machine will carry out activities which may best be described as self-improvement.

²⁵ John McCarthy et al., “A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955,” *AI Magazine* 27.4 (2006): 12-14. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness [Ed. note: yes, they wrote “a some randomness”]. The randomness must be guided by intuition to be efficient.

For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving. [...] The important result that would be looked for would be that the machine would tend to build up within itself an abstract model of the environment in which it is placed. If it were given a problem, it could first explore solutions within the internal abstract model of the environment and then attempt external experiments. Because of this preliminary internal study, these external experiments would appear to be rather clever, and the behavior would have to be regarded as rather “imaginative” (this last paragraph, only John McCarthy’s words in a supplement to the proposal).

That which they did not discuss in their initial proposal? Risks or threats that AI might pose somewhere down the line. And yet such concerns long preceded them. Samuel Butler wrote the following in his 1863 essay *Darwin Among the Machines*:

The upshot is simply a question of time, but that the time will come when the machines will hold the real supremacy over the world and its inhabitants is what no person of a truly philosophic mind can for a moment question.

Before that Henry David Thoreau stated: “A man is rich in proportion to the number of things which he can afford to let alone,” meaning that we could measure wealth and prosperity by the amount of nature we do *not* have to destroy in our quest to prosper. Thoreau himself took much inspiration from Native Americans. As Black Elk put it:

Peace. ...comes within the souls of men when they realize their relationship, their oneness, with the universe and all its powers, and when they realize that at the center of the Universe dwells *Wakan Tanka*, and that this center is everywhere, it is within each of us.²⁶

Yet, as Natives painfully discovered, colonialists had a seemingly insatiable desire for power, money, land, and technological expansion without consideration for the impact such behavior had on animals, plants, and natural environs. Jack D. Forbes referred to this as a type of disease: *wétiko* (a Cree term) which in his usage referred to a form of social cannibalism and collective psychopathy driven by greed that reduced holistic human existence to detached, abstract code (ideas in the service of “progress”).²⁷

²⁶ Jack D. Forbes, *Columbus and Other Cannibals* [revised edition]. New York: Seven Stories Press, [1979] 2008: 188.

²⁷ Ibid: viii. Also, see Russell Means, “Fighting Words for the Future of the Earth,” In *Questioning Technology*, edited by John Zerzan and Alice Carnes, 71-82 (Philadelphia: New Society Publishers, 1991).

Despite entire cultures of skepticism toward extractivist and exploitative economies that pursue technological expansion at any cost, most critics—however extreme they seemed—have typically aimed to regulate and temper developments such as AI rather than abandon or abolish such projects altogether.

Jack Good (born Isidore Jacob Gudak), a colleague of Alan Turing, developed the concept of “intelligence explosion” in 1965 imagining that AI could potentially explode in capacity at hyperspeed once it has acquired basic ingredients such as autonomous processing, self-improvement in key areas, and the ability to design better machines (“Thus the first ultrainelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is curious that this point is made so seldom outside of science fiction. It is sometimes worthwhile to take science fiction seriously”²⁸). Or take the following editorial from Ted Lieu:

Imagine a world where autonomous weapons roam the streets, decisions about your life are made by AI systems that perpetuate societal biases and hackers use AI to launch devastating cyberattacks. This dystopian future may sound like science fiction, but the truth is that without proper regulations for the development and deployment of Artificial Intelligence (AI), it could become a reality. The rapid advancements in AI technology have made it clear that the time to act is now to ensure that AI is used in ways that are safe, ethical and beneficial for society. Failure to do so could lead to a future where the risks of AI far outweigh its benefits.

I didn’t write the above paragraph. It was generated in a few seconds by an AI program called ChatGPT, which is available on the internet. I simply logged into the program and entered the following prompt: “Write an attention grabbing first paragraph of an Op-Ed on why artificial intelligence should be regulated.”²⁹

Ted writes computer code himself and used an AI program to help him write an article warning about the dangers of AI. We see sharper critique in leading thinkers in the field of tech and AI such as Ray Kurzweil (superintelligence advocate) and Bill Joy who wrote the following:

We are suggesting neither that the human race would voluntarily turn power over to the machines nor that the machines would willfully seize power. What we do suggest is that the human race might easily permit itself to drift into a position of such dependence on the machines that it would have no practical choice but to accept all of the machines' decisions. As society and the problems that face it become more and more complex and machines become more and more intelligent, people will let machines make more of their decisions for them, simply because machine-made decisions will bring better results than man-made ones. Eventually a stage may be reached at which the decisions necessary to keep the system running will be so complex that human beings will be incapable of making them intelligently. At that stage the machines will be in effective control. People won’t be able to

²⁸ Raffi Khatchadourian, “The Doomsday Invention,” *New Yorker* 23 November 2015.
<https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom>

²⁹ Ted Lieu, “I’m a Congressman Who Codes. AI Freaks Me Out.” *New York Times* 23 January 2023.
<https://www.nytimes.com/2023/01/23/opinion/ted-lieu-ai-chatgpt-congress.html>

just turn the machines off, because they will be so dependent on them that turning them off would amount to suicide.

On the other hand it is possible that human control over the machines may be retained. In that case the average man may have control over certain private machines of his own, such as his car or his personal computer, but control over large systems of machines will be in the hands of a tiny elite—just as it is today, but with two differences. Due to improved techniques the elite will have greater control over the masses; and because human work will no longer be necessary the masses will be superfluous, a useless burden on the system. If the elite is ruthless they may simply decide to exterminate the mass of humanity. If they are humane they may use propaganda or other psychological or biological techniques to reduce the birth rate until the mass of humanity becomes extinct, leaving the world to the elite. Or, if the elite consists of soft-hearted liberals, they may decide to play the role of good shepherds to the rest of the human race. They will see to it that everyone's physical needs are satisfied, that all children are raised under psychologically hygienic conditions, that everyone has a wholesome hobby to keep him busy, and that anyone who may become dissatisfied undergoes "treatment" to cure his "problem." Of course, life will be so purposeless that people will have to be biologically or psychologically engineered either to remove their need for the power process or make them "sublimate" their drive for power into some harmless hobby. These engineered human beings may be happy in such a society, but they will most certainly not be free. They will have been reduced to the status of domestic animals.³⁰

Unlike Ted Lieu who tricked readers by citing AI, Ray and Bill tricked readers by drawing the entire quote above from the text *Industrial Society and its Future* by Ted Kaczynski aka The Unabomber. Both authors, however, not only condemned Ted Kaczynski's methods (terrorist bomb attacks), they also disagreed with his conclusion (anarcho-primitivism). Ray and Bill both support the development of AI—just "nicer" or more regulated versions. Yet, already by 2009 (less than a decade after Bill Joy quoted the Unabomber), a conference of the Association for the Advancement of Artificial Intelligence (AAAI) declared that, while superintelligence seemed unlikely to ever take place, "Intelligent systems that we barely understand but depend on critically are here (Internet, financial transaction systems, power grid,...)," observed that some robots had acquired the capability of finding their own power source, and described computer viruses as "cockroaches of the autonomous agent world."³¹ GPT-4 managed to trick a human into passing a CAPTCHA test for it ("OpenAI's own tests showed that GPT-4 could deliberately lie to a human worker ('No, I'm not a robot. I have a vision impairment that makes it hard for me to see the images') in order to get help solving a captcha test designed to block non-humans."³²

³⁰ Bill Joy, "Why the Future Doesn't Need," *Wired*. 1 April 2000. <https://www.wired.com/2000/04/joy-2/>

³¹ Asilomar Study on Long-Term AI Futures, 2009. http://erichorvitz.com/AAAI_Asilomar_Study.pdf ; also see John Markoff, "Scientists Worry Machines May Outsmart Man," *New York Times*. 25 July 2009. <https://www.nytimes.com/2009/07/26/science/26robot.html>

³² Stuart Russell, "AI has much to offer humanity. It could also wreak terrible harm. It must be controlled," *The Guardian* 2 April 2023. <https://www.theguardian.com/commentisfree/2023/apr/02/ai-much-to-offer-humanity-could-wreak-terrible-harm-must-be-controlled>

Most skepticism toward the idea that AI could lead to human extinctions stems from those such as Noam Chomsky, AI pioneer Edward Feigenbaum, or roboticist Rodney Brooks who don't see superintelligence anywhere close to development ("It is ridiculous to talk about such things so early— A.I. is eons away," in Edward's words and "I think it is a mistake to be worrying about us developing malevolent AI anytime in the next few hundred years," as Rodney put it).

Yet, as others have already pointed out, AI could kill us even if we do not come close to AGI or superintelligence.

...if artificial intelligence might not be tantamount to "summoning the demon" (as Elon Musk colorfully described it), AI-enhanced technologies might still be extremely dangerous due to their potential for amplifying human stupidity. The AIs of the foreseeable future need not think or create to sow mass unemployment, or enable new weapons technologies that undermine precarious strategic balances. Nor does artificial intelligence need to be smarter than humans to threaten our survival—all it needs to do is make the technologies behind familiar 20th-century existential threats faster, cheaper, and more deadly.³³

In a recent *New York Times* article, Yuval Harari, Tristan Harris and Aza Raskin wrote:

Imagine that as you are boarding an airplane, half the engineers who built the plane tell you there is a 10 percent chance the plane will crash, killing you and everyone else onboard. Would you still board?

In 2022, over 700 top academics and researchers behind the leading artificial intelligence companies were asked in a survey about future AI risk. Half of those surveyed stated that there was a 10 percent or greater chance of human extinction (or similarly permanent and severe disempowerment) from future AI systems. Technology companies building today's large language models are caught in a race to put all of humanity on that plane. [...]

Social media was the "first contact" between AI and humanity, and humanity lost. "First contact" has given us the bitter taste of things to come. [...]

Large language models is our "second contact" with AI. We cannot afford to lose again. But on what basis should we believe humanity is capable of aligning these new forms of AI to our benefit? If we continue with business as usual, the new AI capacities will again be used to gain profit and power, even if it inadvertently destroys the foundations of our society. [...]

AI indeed has the potential to help us defeat cancer, discover lifesaving drugs, and invent solutions for our climate and energy crises. There are innumerable other benefits we cannot

³³ Edward Moore Geist, "Is artificial intelligence really an existential threat to humanity?" *Bulletin of the Atomic Scientists* 9 August 2015. <https://thebulletin.org/2015/08/is-artificial-intelligence-really-an-existential-threat-to-humanity/>

begin to imagine. But it doesn't matter how high the skyscraper of benefits AI assembles if the foundation collapses.³⁴

Yes, a one in ten chance of annihilating our entire world and yet the same AI researchers still work on it. (Compare this risk to Covid's initial 1 in 100 chance of killing an infected person). One may rightly wonder after reading about how virtually all AI researchers acknowledge this major threat to humanity from at least some form or aspect of AI, why would they continue to build it?

Geoffrey Hinton (Google Brain scientist): "I think political systems will use [AI] to terrorize people."

Nick Bostrom (AI philosopher): "Then why are you doing the research?"

Gregory: "I could give you the usual arguments. But the truth is that the prospect of discovery is too sweet. When you see something that is technically sweet, you go ahead and do it, and you argue about what to do about it only after you have had your technical success."³⁵

Amidst all of the discussion about AI, we seem to miss a huge part of the equation: *wétiko*. By seeking artificial intelligence in a society that has made "intelligence" compatible with social cannibalism and psychopathic reasoning, industry creates machines not in the image of humanity (things that can *feel*, develop *empathy*, *suffer*, *celebrate*, and *struggle*) but in the image of themselves: cold, calculating devices that process information toward the acquisition of certain goals. In our world, profit serves as the goal of industry and we have no reason to suspect that AI would not *primarily* serve as a device to fulfill those ends. At whatever cost to the rest of us and our Earth.

One of the creators of xenobots, Michael Levin, said: "To be worried about xenobots, in a world where we already have both natural and human-designed pathogens, is just nuts."³⁶ This "logic" seems fascinating. It sounds as logical as saying: "To be worried about nuclear weapons, in a world where we already have both chemical and conventional weapons, is just nuts." In other words, we do not know now whether xenobots will prove as harmless as monarch butterflies or as dangerous as nuclear weapons but the presence of other harmful phenomena in the world does not ensure its safety by any means. Precisely this type of "genius in one sense, retarded in another" combination seems to permeate the industry. Because industry promotes that type of nonsensical thinking by paying people to pursue narrow goals while necessarily blinding themselves to its potential consequences and/or inculcating a state of cognitive dissonance denial so that they can fall asleep at night.

³⁴ Yuval Harari, Tristan Harris and Aza Raskin "If We Don't Master AI, It Will Master Us." *New York Times*. 24 March 2023. <https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html>

³⁵ Raffi Khatchadourian, "The Doomsday Invention," *New Yorker* 23 November 2015. <https://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom>

³⁶ Joshua Sokol, "Meet the Xenobots, Virtual Creatures Brought to Life," *New York Times*. 3 April 2020. <https://www.nytimes.com/2020/04/03/science/xenobots-robots-frogs-xenopus.html>

Similarly, ChatGPT has turned lots of tables and caused some people to go into panic mode, others to conceptually convulse into open orgasms of utopia, and others to angrily dismiss AI as either useless, irrelevant, or a sign of how far away doomsday or paradise really seem. In response, OpenAI's president, Greg Brockman, said "If you're equally upsetting both extreme sides, then you're doing something right."³⁷ Again, a logical fallacy. Jeffrey Epstein could have said the same thing. After all, revolutionary communists and neo-Nazis both tend to agree in their opposition to human trafficking and sex with children. Flat-earthers, Jehovah's Witnesses, and dogmatic atheists tend to agree on that too. The fact that a person says or does something that revolts other people on "both extreme sides" has actually no bearing whatsoever on a statement or behavior's truthfulness or ethical position. How can such presumably intelligent people (the ones creating "artificial intelligence") spout such obvious fallacies that a ten-year old could see the hole their argument? Probably because such logical bypasses enable them to continue justifying the unjustifiable and the fact that they pursue their mission for profit or curiosity or fame or whatever, but not because they reliably believe it will not threaten all of our lives. In the recent *New York Times* interview with Sam Altman, CEO of OpenAI, they wrote: "To spend time with Mr. Altman is to understand that Silicon Valley will push this technology forward even though it is not quite sure what the implications will be," and went on to quote him paraphrasing J. Robert Oppenheimer (who played a key role in developing the atomic bombs dropped on Nagasaki and Hiroshima): "Technology happens because it is possible."³⁸

In another interview, Sam said, "I think I'm really disconnected from, like, the reality of life for most people."³⁹ Subsequently, he tries to remedy that: "So I really want to go and talk with our users in a lot of different spaces." And yet, it does not seem to occur to him (or perhaps matter) that those users too may not connect well to the lives of most people.

Who wants to believe they contributed to mass extermination of our (or any) species? And yet, E.O. Wilson and others have estimated that industrial civilization leads to the death of approximately 100 species per day. The clock ticks away. And the battery fueling that clock contributes to the problem while society ignores the deaths behind each ticking.

³⁷ Cade Metz, "The ChatGPT King Isn't Worried, but He Knows You Might Be," *New York Times* 31 March 2023. <https://www.nytimes.com/2023/03/31/technology/sam-altman-open-ai-chatgpt.html>

³⁸ Ibid.

³⁹ Lex Fridman. "Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI," Lex Fridman podcast 25 March 2023. https://www.youtube.com/watch?v=L_Guz73e6fw

Imagine Imminence

There are a few people who believe that there is a fairly high-percentage chance that a generalized AI will happen in the next 10 years. But the way I look at it is that in order for that to happen, we're going to need a dozen or two different breakthroughs. So you can monitor when you think these breakthroughs will happen. And you just have to have somebody close to the power cord. [Laughs.] Right when you see it about to happen, you gotta yank that electricity out of the wall, man.

- Barack Obama

The brain of the village idiot and the brain of a scientific genius are almost identical. So we might very well see relatively slow and incremental progress that doesn't really raise any alarm bells until we are just one step away from something that is radically superintelligent.

- Nick Bostrom

What is important is to spread confusion, not eliminate it.

-Salvador Dali

The separation between human and machine is going to blur in a very fundamental way.

- A senior I.B.M. executive

Imagine your life as a pregnant woman, nine months of carrying a living creature, nurturing it, feeding it, until it enters the world and yet, upon birth, you see not a human child but a large mechanical insect. You feel a sense of immediate horror. But that moment does not last because it eats you in an instant. In some sense, this scenario may describe humanity's collective predicament with AI: giving birth to a type of machinery that annihilates the body that produced it. Just as the line between mother and child seems very blurry in the early stages when the fertilized egg seems entirely dependent on the mother (not so now with the advent of mechanical wombs but that actually strengthens the broader point about how far the AI "fetus" has developed), so too does the line between "AI" and "human" blur. We seem to function as a complex cyborgian system building itself with "human" cells currently performing the work done increasingly by mechanical cells until the point at which the mechanical cells play such key and central roles that they can independently (a) acquire their own sources of energy, (b) procreate, and (c) repair, program, and evolve their abilities and systemic structures.

And yet, the metaphor falls short in other senses because, unlike a mother who may feel attached to but not dependent upon the child in her womb, humans have grown increasingly dependent on AI.

And, obviously, the sheer mechanics of billions of people using AI on their cell phones or computers while a few thousand people develop AI from within the respective confines of various corporate domains differs dramatically from a single human bearing a

child in her womb. Yet, despite these distinctions, the metaphor seems useful, perhaps necessary, to illustrate the sense of collective fate we share. And, even if we do not belong to the minority number of “brain cells” of corporations, governments, and AI researchers choosing to proceed or abort the process of AI fetal development, we all belong to the human body subject to the new life we witness upon its birth.

And no, we do not have a guarantee that the “mechanical insect” of AI would kill us immediately upon birth. If we did, I would not get built. But we know that the chance remains. And why do we seem so willing to play with the fire when we could, if we shared resources and power responsibly and equitably, manage to live healthy and happy lives without it?

It leads me to wonder: what happens when we look around at actual insects? What happens inside of us when we observe and interact with dogs and lizards, cats and ferrets, horseshoe crabs and ants, cows and apes in the world? If we really look at them and see how they relate to each other, to us, and to their habitat, how, I wonder, could we possibly believe that we need AI?

Rather than “intelligence” without either body or empathy, we can engage our own bodies and capacities to empathize as the only actual intelligence we need to figure out how to harmonize our lives with the world in which we live.

Paths to Apocalypse: How Many Ways Could We Die By AI?

Apocalypse has a dual meaning: *absolute catastrophe of global proportion* (in popular usage) and *revelation* (Greek source to the English word). This seems very appropriate here in the sense that the catastrophe has now begun to reveal itself. If anyone doubted what *might* lie on the horizon, recent socio-technological developments have given a very clear indication of possibilities.

This section therefore presents a variety of possible means by which AI could trigger, invoke, or otherwise lead to apocalypse (in the sense of absolute catastrophe).

The following scenarios appear in order of likelihood (note that none of these mutually exclude others so one may witness various blends take place, and some of them have already taken place to some degree or another):

(1) Collective Insanity

This scenario has already taken place although in ways not obvious to most of us. We might usefully distinguish between three major shifts in condition. The only general state of existence for humans prior to collective insanity took place in our animal state (probably pre-ape stage). The primordial stage of collective insanity shows itself in early developmental level among chimpanzees wherein a dominant male chimp might, for example, “fight off” a thunderstorm by stomping around, beating his chest, and tearing branches off of trees. Other chimps may feel some degree of consolation and, to that extent, they have gone insane—much like how people feel consoled by patriotism or prayers or military parades. Collective insanity. So we can’t blame AI for making humans go crazy. But AI seems slated to take it to a whole new level and type. If we think of human history as taking place in three stages in this regard, it might look like this:

Stage 1, Moderate collective schizophrenia: Human societies privileged shamans, rainmakers, and medicine folk who sometimes knew their shit (local herbs and their medicinal properties) and sometimes manifested what we might otherwise describe as schizophrenia or psychosis (meaning that they heard, saw, felt, or thought things that other people couldn’t or didn’t). That type of insanity could get rewarded and sometimes even guided the entire group. These same human societies punished and ostracized people who behaved in anti-social manners. If a person behaved in a ruthless way, in an arrogant manner with no consideration for other people, without empathy or ethics then that person could risk losing their position in society altogether.

Stage 2, Structural psychopathy: Since the spread of colonialism in the 1600 and 1700s, we have lived in a world increasingly dominated by another sort of insanity: a world where might makes right. In this world that we live in today, societal treatments of variants of insanity have reversed. People who experience psychosis or schizophrenia now get locked up in institutions or live as outcasts. In contrast, people who exhibit no sense of empathy or ethics (in other words, psychopaths and sociopaths), may get rewarded with top positions managing big banks, leading major corporations, or running the countries we live in. And here in this society, AI gets produced by and for people to serve the system within which it appears: a society that privileges psychopathic behavior.

Now, AI paves the way for a new level of collective insanity...

Stage 3, Deep Insanity: With the rise of a profit-driven “information society” that tears apart communities, concentrates even more power into moneyed elites, and ravages ecosystems, we see a blend of (a) society behaving psychopathically while (b) acquiring a distinct and semi-permanent state of psychosis to the extent that people live in ways completely detached from the animals, plants, and ecologies that support them and, increasingly, even detached from one another. Instead of embodied human community, people in an AI-driven world, find their psychic “center” (point of reference, source of information, ethical guidance, daily news, personal advice, etc.) in a machine without blood or sensation: just a cold, calculating quest for goal fulfillment. If the shaman-centered society gave way to the psychopath-centered society, then the psychopath-centered society now seems on the verge of giving way to a machine-centered society.

Against this background, we might think about the words of Jaron Lanier (who coined the term “virtual reality”):

People survive by passing information between themselves. We’re putting that fundamental quality of humanness through a process with an inherent incentive for corruption and degradation. [...] I still think extinction is on the table as an outcome. Not necessarily, but it’s a fundamental drama. If we can coordinate ourselves to solve the climate crisis it’s a fundamental sign we haven’t become completely dysfunctional. [...] As we enter an era where nothing means anything because it’s all just about power, intermediation and influence, it’s very hard to put ideas out and very easy for them to come across not as intended. [...] The most profound problem here is, can we be sane enough to communicate and coordinate for our survival. [...] Even people who are willing to cooperate may not be able to because they’re not operating in an environment where they’re heard in the ways they imagine. Right now we have no confidence that what we say will be heard correctly. [...] ⁴⁰

From my perspective, the danger isn’t that a new alien entity will speak through our technology and take over and destroy us. To me the danger is that we’ll use our technology to become mutually unintelligible or to become insane if you like, in a way that we aren’t acting with enough understanding and self-interest to survive, and we die through insanity, essentially. ⁴¹

Large Language Models exacerbate our existing insanity by distilling information for a specific source: the Internet. This means that from the get-go, LLMs privilege civilizational societies (those that have acquiesced to psychopath-friendly cultures or already embody them): English, French, Mandarin, Spanish, Arabic, Italian, and so on. About 6,000 languages exist in the world, but the world of the Internet gives disproportionate space to European colonialist languages and other imperial cultures such as Japanese, Chinese, Turkish, Farsi, and Arabic. Indigenous languages and their way of thinking—already detached from their natural habitats by virtue of the Internet’s *de facto*

⁴⁰ Edward Helmore, “‘Extinction is on the table’: Jaron Lanier warns of tech’s existential threat to humanity,” *The Guardian* 27 November 2022. <https://www.theguardian.com/technology/2022/nov/27/jaron-lanier-tech-threat-humanity-twitter-social-media>

⁴¹ Simon Hattenstone. “Tech guru Jaron Lanier: ‘The danger isn’t that AI destroys us. It’s that it drives us insane’.” *The Guardian* 23 March 2023. <https://www.theguardian.com/technology/2023/mar/23/tech-guru-jaron-lanier-the-danger-isnt-that-ai-destroys-us-its-that-it-drives-us-insane>

detachment from nature—get virtually wiped out (again). This means that current AI models amplify existing racist exclusion but, not just racist, these models amplify the types of things people would say when embedded in the mindset of imperial, colonialist, and psychopath-friendly cultures. As Irina Raicu (Santa Clara University), said of the LLM type of AI: “It’s giving you what the collective hive mind on the internet would say.”⁴² If AI functions as a mirror to society, the face in front of the mirror seems like the face of kidnappers doing a promotional video for their product produced in secret by the people they kidnapped—people never seen on the screen whose voice doesn’t matter, just their labor and their use as a product to make the video possible through their suffering.

The big tech companies have spent 20 years harvesting vast amounts of data from culture and everyday life, and building vast, energy-hungry data centres filled with ever more powerful computers to churn through it. What were once creaky old neural networks have become super-powered, and the gush of AI we’re seeing is the result. [...] AI image generators will do what all previous technologies have done, but they will also go further. They will reproduce the biases and prejudices of those who create them, like the webcams that only recognise white faces, or the predictive policing systems that lay siege to low-income neighbourhoods. [...] It is very good at producing what sounds like sense, and best of all at producing cliché and banality, which has composed the majority of its diet, but it remains incapable of relating meaningfully to the world as it actually is. [...] The belief in this kind of AI as actually knowledgeable or meaningful is actively dangerous. It risks poisoning the well of collective thought, and of our ability to think at all. If, as is being proposed by technology companies, the results of ChatGPT queries will be provided as answers to those seeking knowledge online, and if, as has been proposed by some commentators, ChatGPT is used in the classroom as a teaching aide, then its hallucinations will enter the permanent record, effectively coming between us and more legitimate, testable sources of information, until the line between the two is so blurred as to be invisible. [...] AI technologies are bad for the planet too. Training a single AI model—according to research published in 2019—might emit the equivalent of more than 284 tonnes of carbon dioxide, which is nearly five times as much as the entire lifetime of the average American car, including its manufacture. These emissions are expected to grow by nearly 50% over the next five years, all while the planet continues to heat up, acidifying the oceans, igniting wildfires, throwing up superstorms and driving species to extinction. It’s hard to think of anything more utterly stupid than artificial intelligence, as it is practised in the current era.⁴³

Of course, behind these generalizations some persons (as well as some societies) suffer more insanity than others. We can, however, minimize insanity and maximize empathy. Yet, through our newly acquired mechanical deep insanity, we may witness the dawn of a purely logical sounding rationale for doing countless things that *on the surface* may seem very sane yet, precisely because we have a machine as our shaman, as our boss, as our oracle, as our president—rather than a circle of community with fellow forms of life—that provides the basis for a collective insanity taken to a whole new level.

⁴² Alyson Krueger. “We Need to Talk (Just as Soon as I Consult ChatGPT).” *New York Times*. 24 March 2023. <https://www.nytimes.com/2023/03/24/style/ai-chatgpt-advice-relationships.html>

⁴³ James Bridle, “The Stupidity of AI,” *The Guardian*. 16 March 2023.

<https://www.theguardian.com/technology/2023/mar/16/the-stupidity-of-ai-artificial-intelligence-dall-e-chatgpt>

(2) The Mass Mash-up

I first experienced the concept of mash-up in the 1990s with the track “Rebel without a Pause (Whipped Cream Mix)” by Evolution Control Committee that cleverly managed to “mash up” the vocals of rap group Public Enemy with the instrumental music of Herb Albert. That eventually kicked off a trend of mashing together different songs together. Yet, the technology to do so remained largely in the hands of tech/music geeks who knew how to merge distinct compositions and sounds into one cohesive piece. Now, AI technology has brought that ability to “mash-up” disparate works together to the masses and, far from only music, anyone can now, with a simple prompt, mash together anything from art to literature to film (integrating AI into 3D printing will enable similar feats in physical constructions). While it seems like a cute, harmless, and fun technological trick, it takes place in tandem with a massive mash-up between machine and human. Without any apparent agent to laugh or cheer at the ingenuity of the process, humans now serve as the product getting mashed up with the machine world itself. Just as musical and literary “mash-up” processes detach and deform human cultural products from their original context and the world of meaning that made them, humans get detached and deformed within a mechanical world alienated from the ecologies, animals, plants, waterways, birdsong, insect buzzes, terrain limitations, and natural processes that birthed us. The detachment from context leads to a detachment from reality or, at least, a sense of trust in what we can or cannot rely on.

In this scenario, the pervasiveness of AI mechanizes human life and culture to such a degree that distinctions blur ever more. Relationally, humans increasingly interface with AI and use AI to interface with others. This means that AI “impersonators” enter human spheres of concern and compassion with increasing “rights” and power. Societally, AI-facilitated decisions and human-determined decisions blur beyond distinction and, even without physical intrusions on individual bodies, the mental merger between AI and humans means that, with humans growing increasingly mechanical and machines growing increasingly human, the two become permanently bound in a way that forces all members of society comply with this merger. Ethically, humans develop “higher” morals in some respects yet do so in mechanical manners. AI exacerbates existing structures of power inequality so this means that even “noble” goals such as sustainability or distributive justice can get implemented in brutal and/or psychopathic manners without consideration for nuance, exceptions to the rule, or any sense of grace and humanity.

Think about this simple fact: modern mainstream human society exists completely dependent on the Internet which, consisting entirely of ones and zeroes, potentially lends itself to complete co-optation by any entity that also consists of ones and zeroes. In other words, AI can duplicate, re-shape, or fraudulently impersonate whatever you see or hear through a computer, whether a film, a news program, a phone call with a family member, or a video conferencing event with work colleagues. Even your personal death no longer belongs to you because someone using AI (or potentially an autonomous AI program) can now use your digital images, audio, and online content to construct a digital version of you (precisely such a service launched in the summer of 2022 by a start-up named StoryFile allows a person to create a digital version of themselves with whom others can

communicate during or after their funeral).⁴⁴ Increasingly from now and into the coming decade ahead, you have no ability to really know whether that which you see on the screen represents that which you think it does. And, at some point, perhaps, you cease to care.

This means that even if individual persons remain purely flesh and blood (which they apparently will not), society as a whole will have crossed the line from mostly-human cyborg to mostly machine-cyborg. Whether humans continue to exist or not seems mute in this scenario because they will function as cells in a larger mechanical and collectively insane society—simple cells programmed to comply for the sake of serving a larger machine body.

On an outward level it seems as if human life continues and no apocalypse ever took place. Yet, under the surface, humans find themselves bound to AI and embedded in mechanical networks that direct them and increasingly deprive them of unmediated human and human-nature contact. Human life independent of AI remains dead, not an option, and, at best, a luxury afforded elites in temporary moments that provide an illusory sense of living outside of the machine.

(3) Eco-ill-logical F-Ai-lure

In this scenario, AI serves to “benefit” the most powerful the most and this means accruing more wealth to elites and selling more plastic and planned obsolescence to the populace. All while top soil, water supplies, biodiversity limits, environmental costs of pollution, and emissions in the atmosphere now give us only a few decades—perhaps only a few years—to avert irrevocable destruction. Some elites have obviously already given up on the prospects for saving Earth from people like them. So they set their sights on Mars and, in the process waste even more resources, cause even more pollution, and contribute to the killing of even more species here on Earth. This, of course leads to increasing human misery and desperation—speeding the process of destruction further. Already in 2020, “2 billion people did not have access to safe drinking water”.⁴⁵ With nearly 100 million refugees in the world in 2021, 30 years later, “3.4 billion people will reside in countries facing catastrophic ecological threats.”⁴⁶ That may well include you.

⁴⁴ See Victor Tangermann, “AI Allows Dead Woman to Talk to People Who Showed Up at Her Funeral,” *Futurism* 19 August 2022. <https://futurism.com/ai-dead-woman-talk-people-funeral>

⁴⁵ Institute for Economics and Peace, “Ecological Threat Register: Ecological Threat Report 2022: Analysing Ecological Threats, Resilience & Peace,” Sydney, October 2022: p. 22. <http://visionofhumanity.org/resources>

⁴⁶ Ibid: p. 4.

(4) AI Police State

As Geoffrey Hinton hinted at earlier, governments, police, and other authoritarian institutions (from sects to businesses, from paramilitaries to drug syndicates) will use AI to their own ends. As with the Chinese “social credit system”⁴⁷

A brain wave reader that can detect lies. Miniaturized cameras that sit inside vape pens and disposable coffee cups. Massive video cameras that zoom in more than a kilometer to capture faces and license plates. At a police conference in Dubai in March, new technologies for the security forces of the future were up for sale. Far from the eyes of the general public, the event provided a rare look at what tools are now available to law enforcement around the world: better and harder-to-detect surveillance, facial recognition software that automatically tracks individuals across cities and computers to break into phones. [...] One tech firm based in the Emirates with ties to the country’s leadership, Presight AI, sells software nearly identical to products that are popular with the Chinese police. At the conference, its software used cameras and A.I. to identify people, store data about their appearance and track their routes as they wandered the event.⁴⁸

Regular people who accept AI as part of their lives, it seems, wander blindly into a future world that seamlessly blends surveillance capitalism with the surveillance state. Most of this happens through cameras we never notice, online surveillance we cannot see, or GPS-tracking of cell phones and computers that lie far beyond our sight or awareness.

(5) Grey Goo

Inspired by (or perhaps depressed by) Kurt Vonnegut’s 1963 novel *Cat’s Cradle*, in which scientists accidentally triggered a “gray-goo-like” disaster called ice-nine that froze the oceans, some tech researchers have used the term “gray-goo” to refer to any unintended consequence of nanotechnology or similar tech that can inadvertently interfere with existing molecular equilibrium to produce something that spreads, replicates, or otherwise disrupts that equilibrium to our catastrophic disadvantage. Bill Joy wrote:

Among the cognoscenti of nanotechnology, this threat has become known as the “gray goo problem.” Though masses of uncontrolled replicators need not be gray or gooey, the term “gray goo” emphasizes that replicators able to obliterate life might be less inspiring than a single species of crabgrass. They might be superior in an evolutionary sense, but this need

⁴⁷ Rachel Botsman, “Big data meets Big Brother as China moves to rate its citizens: The Chinese government plans to launch its Social Credit System in 2020. The aim? To judge the trustworthiness – or otherwise – of its 1.3 billion residents.” *Wired* 21 October 2017. <https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>

⁴⁸ Paul Mozur and Adam Satariano. “A.I., Brain Scans and Cameras: The Spread of Police Surveillance Tech.” *New York Times* 30 March 2023. <https://www.nytimes.com/2023/03/30/technology/police-surveillance-tech-dubai.html>

not make them valuable. The gray goo threat makes one thing perfectly clear: We cannot afford certain kinds of accidents with replicating assemblers. Gray goo would surely be a depressing ending to our human adventure on Earth, far worse than mere fire or ice, and one that could stem from a simple laboratory accident. Oops.

Indeed, it sounds like a sorry fate and AI development could exacerbate such risks tremendously by placing all technical research on hyperspeed.

(6) AI State

In this scenario, the “servant-master” role inverts and autonomous robot and computer network agents coerce humans to overtly perform the duties that sustain mechanical life. In an “AI state,” humans exist in a hostage condition whereby AI would have the power over electric grids and/or energy supply and/or Internet and/or weapons of mass destruction. (Such a scenario would not necessarily require ordinary people to even know about these circumstances, as it would only require decision-makers to know).

(7) The Matrix

Robots take over and relegate human consciousness to simulated realms. This possibility seems least likely and almost certainly in the realm of science fiction but, of course, who knows?

Only these last two scenarios, the ones least likely and the ones so far receiving the most attention, require AGI or superintelligence to take place.

Concluding Thoughts

So yes, as stated in the beginning, I thought we had more time. I thought we had at least a few years before we might enter Stage 3 of our collective insanity. I thought that we would not need to act nearly as fast as it now seems that we would need to (if we want to avert an imminent apocalypse).

Now it feels like we might only have a window of a few months (if even that).

Yet, in that error, we can find opportunity if people in a sudden moment of collective insight halt the development of AI.

I see two windows for human/humane survival.

(1) **Post-Apocalypse.** Nuclear war or other catastrophe aborts the birth of any autonomous A.G.I. agents and—miraculously—a sufficient number of humans survive to learn from what happened.

(2) **Pre-Apocalypse.** If AI directly and indirectly accentuates and exacerbates existing trends toward our ecological extermination, we could avert this by opting out of the very profit-based system that has made AI not only possible but also inevitable.

Even if the risk remains only one in a hundred (not one in ten as many AI researchers believe) and even if our chance of success remains only 1 in a million, placing our efforts on the second option seems to make sense—especially if we can do so in ways that benefit us here and now in other ways (such as sharing wealth and resources, less unnecessary consumption, less stress, more quality time with friends and family, more experiences in real life rather than digital worlds, etc.).

At the end of March 2023, **prominent persons in AI development, academia, and the tech industry** (such as Turing Prize winner Yoshua Bengio, author of *Artificial Intelligence: a Modern Approach* Stuart Russell, Elon Musk, Co-founder of Apple Steve Wozniak, Yuval Noah Harari, Andrew Yang, Skype co-founder Jaan Tallinn, Co-Founder of Pinterest Evan Sharp, inventor of associative neural networks John Hopfield, President of the Bulletin of the Atomic Scientists Rachel Bronson, and Max Tegmark at the MIT Center for Artificial Intelligence & Fundamental Interactions) **all called for a pause in “giant AI” development.**⁴⁹

⁴⁹ Yoshua Bengio et al. “Pause Giant AI Experiments: An Open Letter.” *Future of Life* March 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments>

They wrote:

AI systems with human-competitive intelligence can pose profound risks to society and humanity [...] Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control. [...] *Should* we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders.

This may seem like a huge step. And yet they do not take it nearly far enough. They recognize the threat but they cannot seem to release their own addiction to—and investments in—AI development. So they called only for a short (6 month) pause in research and, even then, not *all* AI research, only those systems more powerful than GPT-4. Indeed, the open letter bears the hollow ring of a country losing an arms race or war and calling for a temporary ceasefire with their opponent in order for them to have time to catch up. OpenAI beat them and, as of March 29, no one from OpenAI signed the open letter.

In response to that open letter, another group of researchers (the ones behind the well-known-to-insiders article “Stochastic parrots”⁵⁰) published their own open letter.⁵¹ Therein, they decried the “fearmongering,” “AI hype,” and “longtermism,” that focuses on hypothetical scenarios involving AGI and superintelligence rather than “the actual harms resulting from the deployment of AI systems today.” They did not depict end-of-the-world-scenarios but did describe three examples of severe risks of AI that the first open letter ignored: “1) worker exploitation and massive data theft to create products that profit a handful of entities, 2) the explosion of synthetic media in the world, which both reproduces systems of oppression and endangers our information ecosystem, and 3) the concentration of power in the hands of a few people which exacerbates social inequities.” And yet, even the authors of this response-letter support AI research and development.

We cannot leave our fate to governments and certainly not any set of tech barons or scientists (those that signed the first letter, those that signed the second, or those that signed neither). When people organize together to radically re-organize society based on respectful human relationships and respect for relationships with animals and plants, *then* we have a chance.

When people exchange without money, share in solidarity, and produce and consume as much as possible locally, *then* we have a chance. When people measure prosperity in terms of health, happiness, and harmony with nature (the eco-systems from which we all come and which we remain dependent upon), *then* we have a chance.

If we have a chance at all, we have it now. We have no reason to wait any longer.

⁵⁰ Timnit Gebru, Emily M. Bender, Angelina McMillan-Major, and Margaret Mitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” *FAccT* 21, 3–10 March 2021: 610–623. <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>

⁵¹ Timnit Gebru, Emily M. Bender, Angelina McMillan-Major, and Margaret Mitchell, “Statement from the listed authors of Stochastic Parrots on the ‘AI pause’ letter,” *The DAIR Institute* 31 March 2023. <https://www.dair-institute.org/blog/letter-statement-March2023>

Glossary

Algorithm: A procedure for solving a problem or, as in computer science, precise technical instructions in a finite sequence in order to perform a specific task. (Contrast with “heuristic”: open-ended, experimental, and/or unspecific means of solving problems).

Alignment: The quest to align AI performance with the goals of their designers (in other words, to restrict AI from developing independent and autonomous trajectories and processes harmful or contrary to designer intent).

Artificial General Intelligence or AGI: A broad term referring to a level or capacity of AI that does not remain confined to specialized areas but can generate a wide range of creative responses in something that closely mimics or rivals human intelligence with a “superintelligence”.

Artificial Intelligence or AI: This refers largely to advanced task-solving machines but also the general prospect of creating the capacities for machines to “think,” process information creatively, analytically, logically, and in ways that meaningfully connect and communicate with the world we live in. Many, if not most, AI researchers dislike the term “AI” but we all seem stuck with it. Some prefer “machine learning” (ML). Depending on how one defines AI, it already exists or maybe never will.

Artificial Neural Networks or ANNs: Inspired by the neural networks that enable brains to function, ANNs use many nodes (or units) that receive and send signals to other nodes in ways that enable the collective network to better align its processes with the instructed goal.

Augmented Reality or AR: Supplementing real life experiences with digital images and experiences often through the use of tools such as goggles, earbuds, or cellphones that reflect both the actual world as well as transformations of that world by digital interventions. Google Street View or Pokémon Go serve as examples of AR. More or less synonymous with Mixed Reality (MR), both of which, along with Virtual Reality, fall under the umbrella term Extended Reality (XR). (XR, in other contexts, stands for Extinction Rebellion).

Black box: The term used to describe the phenomena that AI programs have reached such as degree of complexity (for example, 175 billion “parameters” or digital “neurons” in a recent chatbot) that scientists and programmers cannot ascertain how the AI program reached its conclusion. This can lead to unforeseen consequences. For example, according to one study, an AI program could identify a person’s sexual orientation with 81% accuracy (better than human observers) and yet researchers could not determine the means by which the AI program arrived at its conclusions.⁵²

ChatGPT: OpenAI’s chat program released to the public to much fanfare in November 2022. Its release constituted a watershed moment when people realized how they could use to program to produce shockingly well-written essays, pass tests, answer all sorts of questions, and even write computer code.

⁵² BBC “Row over AI that ‘identifies gay faces’,” *BBC* 11 September 2017.

<https://www.bbc.com/news/technology-41188560> ; Yilun Wang and Michal Kosinski, “Deep neural networks are more accurate than humans at detecting sexual orientation from facial images,” *Journal of Personality and Social Psychology*, 114(2), 246–257. <https://osf.io/zn79k/>

DALL-E: The name of a visual art AI program that creates images based on “prompts” (the user writes an instruction such as “paint a picture of Uncle Sam barfing a rainbow on a dinosaur” and it creates the image). The name itself combines references to WALL·E, a robot in a mainstream movie and Salvador Dalí (1904-1989), pioneering surrealist painter. Numerous other AI art programs have followed.

DeepFake: This refers to fabricated yet realistic digital images, audio, and videos of people that make it seem as if a person said or did something that they never did.

Deep Learning or DL: This refers to a type of Machine Learning algorithms that employs multiple layers of processing advanced features from raw input.

Generative AI: Stuff that makes stuff. AI that makes music, images, speech, or text.

Generative Adversarial Networks or GAN: A type of Deep Learning model used today primarily to create new images, video, and audio through internally competing systems. GAN basically consists of two neural networks: one that categorizes and organizes data (the *discriminator*) while the other one (the *generator*) creates novel material. One such GAN (on ThisPersonDoesNotExist.Com) generated a “new” human face every two seconds by compiling different images on the Internet into a “new” composition. These “new” or “nonexistent” people quickly fed the Internet with images for fake online account profiles.

Generative Pre-trained Transformer 4 or GPT-4: The Large Language Model (LLM) by OpenAI that serves as the full AI version behind the more limited ChatGPT. Released in March 2023.

Grey Goo Problem: Nanotechnologists use the term “grey goo problem” (derived from Kurt Vonnegut’s *Cat’s Cradle*), to describe the risk of self-replicating nanobots that replicate so much that they destroy certain ecologies or all life on Earth.

Hallucinations: A term currently used to describe AI errors.

Large Language Models or LLM: A type of narrow AI focused on generating predictive text to match patterns that it has previously detected. With millions of images, audio files, and texts drawn from the Internet, LLMs detect patterns that seem to match its instructed goals and uses those to predict a plausible response to any given question or task. This means that LLM often may give a “likely” answer rather than a “true” one because it reflects the human-made data from which it draws its predictions.

Maker shop/Maker scene: This refers to the 3D production labs that fabricate computer-designed products and the subculture that has arisen around them.

OpenAI: An AI development company founded in San Francisco in 2015 by Sam Altman, Ilya Sutskever, Elon Musk, Jessica Livingston, Peter Thiel, and others. Financed with billions of dollars from the founders and Microsoft.

Nanobots: Extremely tiny devices roughly one millionth of a meter, ranging in size from 0.1 to 10 micrometers using microscopic or molecular components. To grasp the size: a human hair measures about 80 micrometers wide.

Parameters: Numerical values in LLMs that define each node in their structure.

Prompt engineering: The ability to formulate instructions to AI programs in order to acquire the desired results.

Singularity or the “**technological singularity**”: In mathematics and mechanics, the term refers to the point at which a mathematical function behaves chaotically, undefined or nondeterministically, or at which one cannot predict mechanical behavior. In relation to AI, singularity generally refers to a hypothetical point in the future in which technological development reaches a point of no return characterized by machine autonomy, absolute human dependency, and/or the domination of humans by an AI “superintelligence”.

Time to Edit or **TTE**: One means of measuring technological approach toward singularity. According to this idea, the main problem with AI language production involves the need to post-edit whatever text an AI system produces. If it took 320 minutes to edit five pages or 2,500 words of AI text ten years ago, 160 minutes to edit five pages five years ago, and 80 minutes to edit five pages of AI-produced text today, that would both show a reduction of 50% TTE every five years as well as provide a general guide to expect a continual halving approximately every five years until “singularity” arrives in about 5 years or whenever TTE reaches “zero” (according to the creators of TTE, the “zero” point amounts to 1 second per word for a professional translator/editor or 40 minutes for a five page document of 2,500 words).⁵³

Transhumanism: The ideology that advocates a merging of human and machine (via robots transforming into something indistinguishably human online or offline, and/or humans transforming into something distinctly mechanical).

Turing Test: Formulated by computer pioneer Alan Turing (1912-1954), the Turing Test began as a means to replace the unanswerable question “can machines think?” with an answerable question, “can a machine communicate in ways indistinguishable from that of a human?” In the latter case, one can conduct simple experiments to determine success or failure. Many people have, however, regarded the Turing Test as a means of affirming *de facto* consciousness in machines (e.g., if you cannot tell whether you communicate with a robot or a human then you might as well treat the writer—human or robot—as a *de facto* human). With the release of humanlike chatbots, the Turing Test now seems outdated and obsolete as a meaningful metric. Today, “reverse Turing Tests,” such as CAPTCHA and reCAPTCHA, routinely take place online by bots that make users like you pass tests to register as “human”.

Virtual Reality or **VR**: This refers to a digital system that provides visual and/or audio or other sensory stimuli to simulate real world experience.

Xenobots: Synthetic and programmable organisms of animated biological material—a bit like mini-Frankenstein monsters. The potential for anyone to use xenobots in water supplies or to inject in animal or person’s blood to produce a certain effect remains an open and eerie question—especially if such biological machines integrate with autonomous AI systems.

Arc of Justice Angel Press N© 2023 (revised versions 12 June 2023 and 24 June 2025).

⁵³ See Translated. “This Is the Speed at Which We Are Approaching Singularity in AI.” <https://translated.com/speed-to-singularity> and Darren Orf, “Humanity May Reach Singularity Within Just 7 Years, Trend Shows,” *Popular Mechanics* 23 January 2023. <https://www.popularmechanics.com/technology/robots/a42612745/singularity-when-will-it-happen/>